DOCUMENT RESUME

ED 199 655 CS 005 946

AUTHOR Fry, Edward
TITLE A Partial Reading Model Utilizing Language Unit Size by Frequency.
PUB DATE Dec 80
NOTE 11p.; Paper presented at the Annual Meeting of the National Reading Conference (30th, San Diego, CA, December 3-6, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Hypothesis Testing; *Models; *Reading Processes; *Reading Research; *Research Problems; Research Utilization; *Word Frequency; Word Lists
IDENTIFIERS *Linguistic Units

ABSTRACT

Modern reading theories that are concerned with the reading process of either mature or immature readers show no clearcut concern with unit size. The various graphic units that are encountered by the reader include letters, graphemes, clusters (such as blends), syllables, phonograms, affixes and roots, words, phrases, sentences, paragraphs, chapters, and whole stories. The frequency of graphic units has a demonstrable effect on the reading process. There is little doubt that frequency does not play a prominent part in most theoretical reading models; however, frequency can tell something important about the reader's use of different language units. While this is only a part of a reading theory, any reading theory that omits any consideration of frequency of unit sizes is in itself a partial theory. (HOD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED199655

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Edward Fry
Director, Reading Center
Rutgers University
New Brunswick, New Jersey 08903

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Edward Fry

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

A PARTIAL READING MODEL UTILIZING LANGUAGE UNIT SIZE BY FREQUENCY

Edward Fry
Rutgers University

The unit size problem has been around for some time. Mathews (1966) reports that ancient Greek and Roman children were taught to read by first mastering letters, then syllables, then words and finally sentences. But just because various units have been around for some while doesn't mean that we know which unit to lean upon. Chall in 1967 found a "Great Debate" still going on between phonics, presumable graphemic units, and whole words. Gleitman and Rosen (1973) have rekindled some interest in the syllable as a teaching unit. The syllable has fallen into some disrepute from the once high honor it held when America was an infant nation - both the New England primmer, and the Webster Spellers, emphasized the syllable as a major beginning literacy unit. Even the

CS005946

famous McGuffey readers, used so heavily in the last century, introduced new words in a syllabicated form, though McGuffey also advocated both phonics and the whole word method.

Modern reading series tend to have a goodly dose of phonics laid on top of a whole word beginning reading approach, plus a pious reverence for comprehension at all levels. In the middle of this century the dose of phonics was considerably lighter. The trends on emphasis or de-emphasis of various unit sizes are at best not well substantiated by research results.

Likewise, modern reading theories which are often concerned more with the "reading process" of either mature or immature readers show no clearcut concern with unit size. Often the word is seen as the fundamental unit being processed, or else the fundamental process is really very abstract such as a schema or past experience. Hence the graphic unit is not of great importance. However, some theories do concern themselves with individual letters or even parts of letters, then they infer a growth of unit sizes. LaBerge and Samuels (1974) postulate that smaller units achieve "automaticity" when larger units are used. Venezsky and Calfee (1970) postulate the interest concept of LMU or Largest Manageable Unit being used by the reader.

This paper will not attempt to take sides on the "optimal" sized unit but rather attempt by simple graphic progression to layout the various sizes of graphical units. But its major purpose is not to state the obvious, namely that there

are graphical units of increasing size and complexity, but that most of these units are ameniable to a basic research technique called frequency counts. This is not new information but it is basic information that is often ignored or slighted by both reading curriculum developers and theoretical reading process model builders.

The history of word counts is shrouded in antiquity. It is possible that in 1588 a chap called Timothe Bright utilized one in developing a shorthand system and probably one was used in 1721 in developing an English dictionary (DeRocher et al 1973). In more modern times Ayres (1915) counted 400,000 running words as a background for his Measuring Scale for Ability in Spelling. But all our modern research in word counting is based more upon the work of Thorndike in 1921 which was published in his classic Teachers Word Book of 10,000 Words. The research in words has continued up to relatively recent times with Kucera and Francis' Analysis of Present Day American English in 1967 and John Carroll's American Heritage Word Frequency Book in 1971.

Frequency counts of units smaller than words have also been important. Cryptographers from Edgar Allen Poe to modern times have utilized letter frequency accounts, though unfortunately the designers of the present typewriter keyboard did not take them into account before the present keyboard became so standardized, nor apparently do authors of penmanship books.

Modern phoneme-grapheme counts stem largely from the work of Dr. Paul Hanna at Stanford University who in 1951 guided a student named Thomas Moore in a hand phoneme-grapheme

frequency count of 3000 words. A much larger study by Hanna, Hanna, Hodges, and Rudorf was done with USOE sponsorship in 1967. This study is partly summarized in the book Spelling, Structure and Strategies (Hanna, Hodges, and Hanna 1971). The Hanna and Hodges frequency count of phoneme-grapheme correspondence is comparable in size and importance to the Carroll word count.

A third major frequency count is the syllable count done by Sakiey (1977). Earlier work in syllable counting is at best fragmentary and at worst unobtainable. The best early work was done by Dewey (1923 and 1950). He counted not graphic syllables but phonetic or speech syllables. However some work in counting graphic syllables was done by Dolch (1940) and Osburn (1954) but they were pre-computer studies. Sakiey not only had computers available but a massive background of data, the Carroll count of 5 million running words with which to "weight" her syllable count. Hence should any curriculum builder or reading model constructor care, there is excellent information available to answer the simple question "Which are the most commonly read syllables?"

There are very few frequency counts of reading units larger than a word. Most of the work in this area seems to be in identifying the unit sizes and in determining their importance in the reading process. For example, the work now going on in figurative language, story grammars, schema, and discourse analysis is concerned both with graphic units and with cognitive units. Increasingly there is evidence not only that larger than word units exist, but that they

can be graded; that is that some are easier than others. Frequency counts that objectively determined "commonness" would be highly desirable.

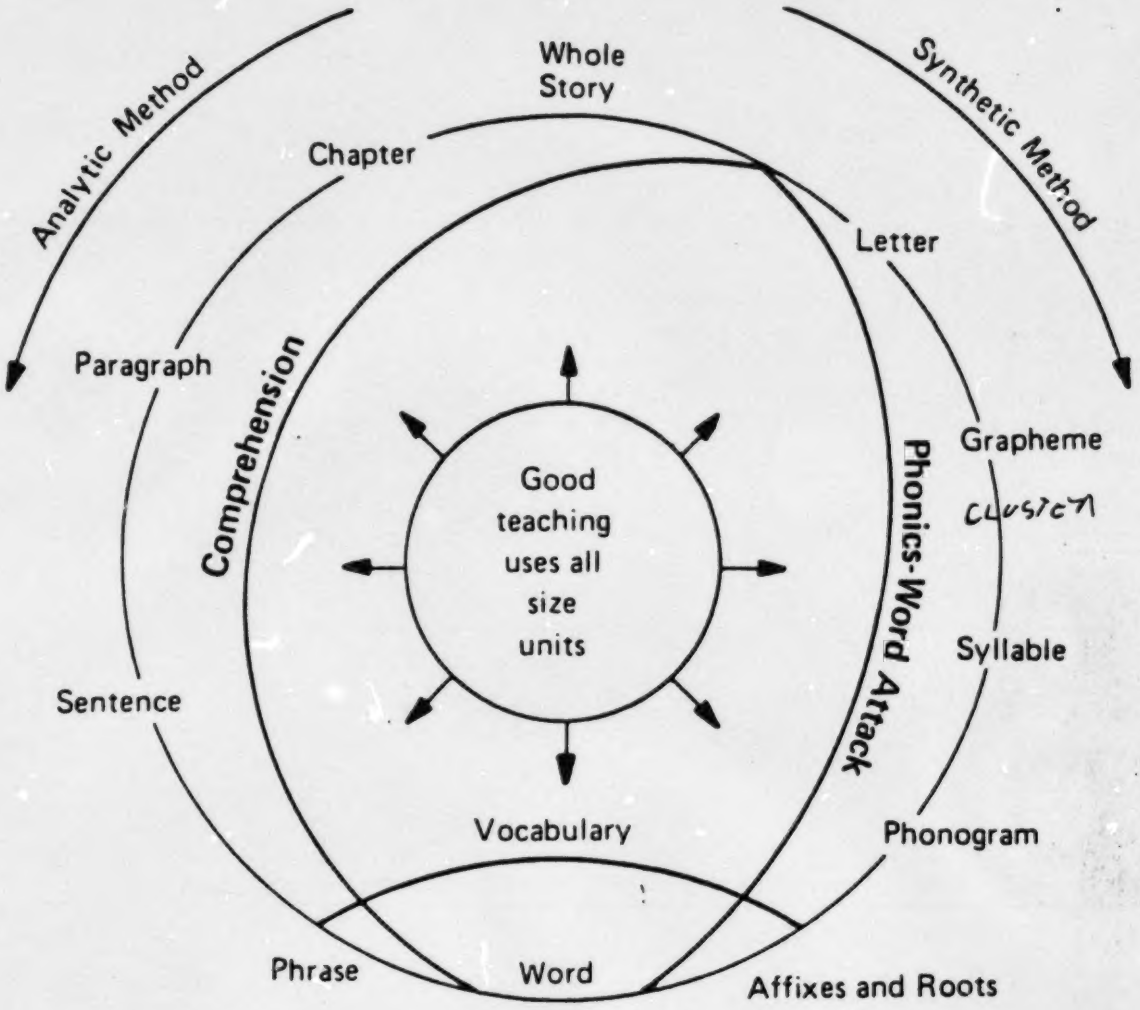
Hopkins and Moe (1978) did have limited success in attempting to count word strings or groups of 2, 3, 4, words utilizing a computer and text from childrens'trade books used in primary grades. They found some commonness for 2 word strings, a little for 3 word strings and almost none for 4 word strings. This research shows that counting of some units larger than words can be done but it is difficult both from data collection and utility conceptualization.

Perhaps for the sake of completeness we should illustrate the various graphic units that are encountered by the reader. They are: Letters, Graphemes, Clusters (such as blends), Syllables, Phonograms, Affixes and Roots, Words, Phrases, Sentences, Paragraphs, Chapters, and Whole Stories. This list could possibly be extended a bit, for example at the lower end below the "letter" we might have some "distinctive features" and at the upper end beyond the "whole story" we might have the trilogy or series of books, but the basic list will serve as a basic core of graphic unit sizes.

(Insert Illustration About Here)

ILLUSTRATION

GRAPHIC UNIT SIZES
USED IN TEACHING READING



Now before some "top down" theorist such as Goodman, or even some interactivist like Rummelhart, attacks the graphic unit as hopelessly "bottom up", or even "bottom out", let me waive a flag of truce and hide behind the title which states that "unit size by frequency" is only a partial theory. Its claim to academic respectability is that it is objective and verifiable, something some other theories are not.

If one wishes to look at either the reading process or learning to read as more, or mostly, cognitive I certainly have no objections. In fact if someone wishes to do a frequency count of schema or story grammar types, I would be most delighted. I would even agonize with them because graphically the idea of doing a frequency count of sentences or paragraphs is difficult not only mechanically but conceptually.

Some of my own work on sentence types showing that certain types of sentences such as those with little "distance" between the subject and verb are easier to read and remember than sentences with greater distance between the subject and the verb tends to show that certain classes of sentences operate in the reading process differently. Hence I could postulate that some sentence types could be counted, but as for objectively counting exact sentences, the prospect boggles the mind if not the largest computer memories.

There is little doubt that frequency of graphic units has a demonstrable effect on the reading process. For example, Finn (1977) found that "cloze easiness" (chance of getting a cloze item correct) correlated with frequency (Carrolls SFI)

at .55. In another study, Steig (1980) using a technique similar to Shannon's Guessing Game as modified by Coleman and Miller (1968) of having students successively guess 100 words of text, found that high frequency words were guessed correctly significantly more often than low frequency words.

There is also little doubt that frequency does not play a prominent part of most theoretical reading models. Singer and Ruddell indexed some 484 content items in their Theoretical Models and Process in Reading, 2nd Edition, and "frequency" was not among them. This is despite the fact that some of the largest and, in my opinion some of the most basic studies related to reading have been in the area of frequency. Perhaps writers for a 3rd edition will not let this omission continue.

"Frequency can tell us something important about the readers use of different language units." While this is only a part of a reading theory, any reading theory which omits any consideration of frequency of unit sizes is in itself a partial theory.

References

Ayres, L. P. A measuring scale for ability in spelling. New York: Russell Sage Foundation, 1915.

Carroll, J. B., Davies, P., & Richman, B. American heritage word frequency book. Boston: Houghton Mifflin, 1972.

Chall, J. Learning to read: the great debate. New York: McGraw-Hill, 1967.

Coleman, E. B. & Miller, G. R. A measure of information gained during prose learning. Reading Research Quarterly, 1968, 3, 369-386.

DeRocher, J. E., Miron, M. S., Patten, S. M. , & Pratt, C. C. The counting of words. Syracuse, New York: Syracuse University Research Corporation, 1973.

Dewey, G. Relativ frequency of English speech sounds. Cambridge: Harvard University Press, 1923.

Dewey, G. Relativ frequency of English speech sounds. Cambridge: Harvard University Press, 1950.

Dolch, E. W. Sight syllables versus letter phonics. Elementary School Journal, 1940, 41, 38-42.

Gleitman, L. R. & Rosin, P. Teaching reading by the use of a syllabary. Reading Research Quarterly, Summer 1973, 8, 447-483.

Finn, P. J. Word frequency, information theory and cloze performance: a transfer feature theory of processing in reading. Reading Research Quarterly, 1978, 13, 508-537.

Hopkins, C. J. & Moe, A. J. "The computer assisted identification of common word strings from the test of children's trade books in primary grades." Paper presented at the meeting of the International Reading Association, Houston, May, 1978.

Kucera, H. & Francis, W. N. Computational analysis of present-day American english. Providence, R.I.: Brown University Press, 1967.

LaBerge, D. & Samuels, S. J. Toward a theory of automatic information processing in reading. Cognitive Psychology, 1974, 6, 293-323.

Mathews, M. M. Teaching to read: historically considered. Chicago: University of Chicago Press, 1966.

Osburn, W. J. Teaching spelling by teaching syllables and root words. Elementary School Journal, 1954, 55, 32-41.

Sakiey, E. H. Syllables: a weighted graphemic inventory. Unpublished dissertation. Rutgers University, New Brunswick, New Jersey, 1977.

Steig, J. Schema theory, word frequency and anticipation in reading. Unpublished dissertation. Rutgers University, New Brunswick, New Jersey, 1980.

Thorndike, E. L. The teacher's word book of 10,000 words. New York: Teachers College Press, Columbia University, 1921.

Venezsky, R. L. & Calfee, R. C. The reading competency model. In Harry Singer and Robert Ruddell (Eds.) Theoretical models and processes of reading. Newark, Delaware: International Reading Association, 1970, 273-291.